JOURNAL OF COMPLEXITY 2, 78–94 (1986)

Strongly Optimal Algorithms and Optimal Information in Estimation Problems*

M. MILANESE, R. TEMPO, AND A. VICINO

Dipartimento di Automatica e Informatica, CENS–CNR, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Torino, Italy

This paper studies some aspects of information-based complexity theory applied to estimation, identification, and prediction problems. Particular emphasis is given to constructive aspects of optimal algorithms and optimal information, taking into account the characteristics of certain types of problems. Special attention is devoted to the investigation of strongly optimal algorithms and optimal information in the linear case. Two main results are obtained for the class of problems considered. First, central algorithms are proved to be strongly optimal. Second, a simple solution is given to a particular case of optimal information, called optimal sampling design, which is of great interest in system and identification theory.

1. INTRODUCTION

One setting of information-based complexity (see Traub and Woźniakowski, 1980; Traub *et al.*, 1983) may be sketched as follows. One is interested in approximating a function $S(f) \in G$ of an element f of a set F (f and S are called respectively problem element and solution operator). The element f is not known exactly but only approximate information is available and is given by $N(f) + \eta = y \in Y$, where N is called information operator and η belongs to a bounded set of Y. An approximation to $S(f)$ can be obtained by acting on y by means of an operator ϕ (called algorithm). By defining a suitable measure of the approximation error, an optimal algorithm is one which minimizes the maximum approximation error for all possible f and η.

Recent papers have shown that a fairly wide class of estimation, identification, and prediction problems, typical in system and control literature, may be embedded in the framework of this theory (Belforte *et al.*, 1982; Milanese and Tempo, 1985).

*This work was partly supported by funds of Ministero della Pubblica Istruzione and of CENS–CNR.

0885-064X/86 $3.00

While the general theory is mainly concerned with the investigation of properties of algorithms and information, estimation contexts generally require the study of constructive aspects of optimal algorithms and optimal information. Feasible algorithms can be obtained by taking advantage of the fact that for these classes of problems particular assumptions may be made: essentially F, G, and Y are finite-dimensional linear spaces; G and Y are equipped with the max-norm; and $\dim F \leq \dim Y$.

Under these conditions, Milanese and Tempo (1985) show that optimal algorithms can be easily derived for linear problems (S and N linear). In this paper we continue the investigation of this class of problems in two directions.

The first direction is to look for optimality concepts stronger than global optimality, which takes into account worst cases of both f and η. If a worst case with respect only to f or to η is considered, locally optimal algorithms can be defined, called y-strongly and f-strongly optimal algorithms. Local optimality has been studied previously in slightly different contexts by Milanese and Belforte (1982) and Traub *et al.* (1983), who give some results on y- and f-strong optimality.

In this paper f- and y-strong optimality conditions are investigated, restricted to the class of "correct" algorithms, i.e., algorithms which map the exact information $N(f)$ on the problem solution $S(f)$, $\forall f \in F$. The concept of correctness of an algorithm makes sense only if the dimension of F is less than or equal to that of Y as supposed here. Actually, almost all of the estimators met in classical estimation theory are correct in this sense.

The main result along this line is that the (globally) optimal central algorithm derived by Milanese and Tempo (1985) is proved to be also f-strongly and y-strongly optimal (Theorem 2).

The second line of investigation is related to the optimal information problem. This consists in looking for the information operator which guarantees the minimum approximation error among all possible information operators of the same cardinality ($\dim Y$). A particular case is investigated, called optimal sampling problem in the identification context (Goodwin and Payne, 1977), where problem elements f are supposed functions of time and information is restricted to sampling operations. In this case an optimal information operator (in both a global and a local sense) can be easily computed. Moreover, using the derived optimal sampling times, a linear strongly optimal algorithm can be obtained (Theorems 5 and 6).

2. Definitions and Notations

Information-based complexity is concerned with the approximation of a given transformation S of an unknown problem element f, using knowledge of the set of all possible problem elements and of a certain number of measurements of f, possibly corrupted by noise. Formally, let F be a linear space over the real field and let F_0 be a subset of F. Consider an assigned operator S (possibly nonlinear), called solution operator, mapping F into G,

$$S: F \to G,$$

where G is a linear normed space over the real field. The goal is to approximate a solution element $S(f) \in G$, with $f \in F_0$ having only limited information available on f.

Let us define a (possibly nonlinear) operator N, called information operator, mapping F into a linear normed space Y:

$$N: F \to Y.$$

In general, for any f belonging to F_0, $N(f)$ may be considered known not exactly, but only with some error η:

$$y = N(f) + \eta.$$

The error η is assumed unknown but bounded by a fixed quantity $\rho \geq 0$:

$$\|N(f) - y\| = \|\eta\| \leq \rho. \tag{1}$$

An algorithm ϕ is an operator, in general nonlinear, mapping Y into G:

$$\phi: Y \to G.$$

A geometric sketch, showing the spaces and operators introduced above, is shown in Fig. 1.

Let us define three sets in the spaces F, Y, and G which play a fundamental role in the development of the theory:

$$E_F(y, N, \rho) = \{f \in F_0 : \|N(f) - y\| \leq \rho\} \tag{2}$$

$$E_Y(f, N, \rho) = \{y \in Y : \|N(f) - y\| \leq \rho\} \tag{3}$$

$$E_G(y, N, \rho) = S\{E_F(y, N, \rho)\}. \tag{4}$$

In the above definition (2), we consider only approximate information y belonging to a subset $Y_0 \subset Y$ such that

$$Y_0 = \{y \in Y : E_F(y, N, \rho) \neq \emptyset\}. \tag{5}$$

Furthermore, we assume that the sets defined in (2) and (4) are bounded; in fact, this assumption is always satisfied in well-posed estimation problems.

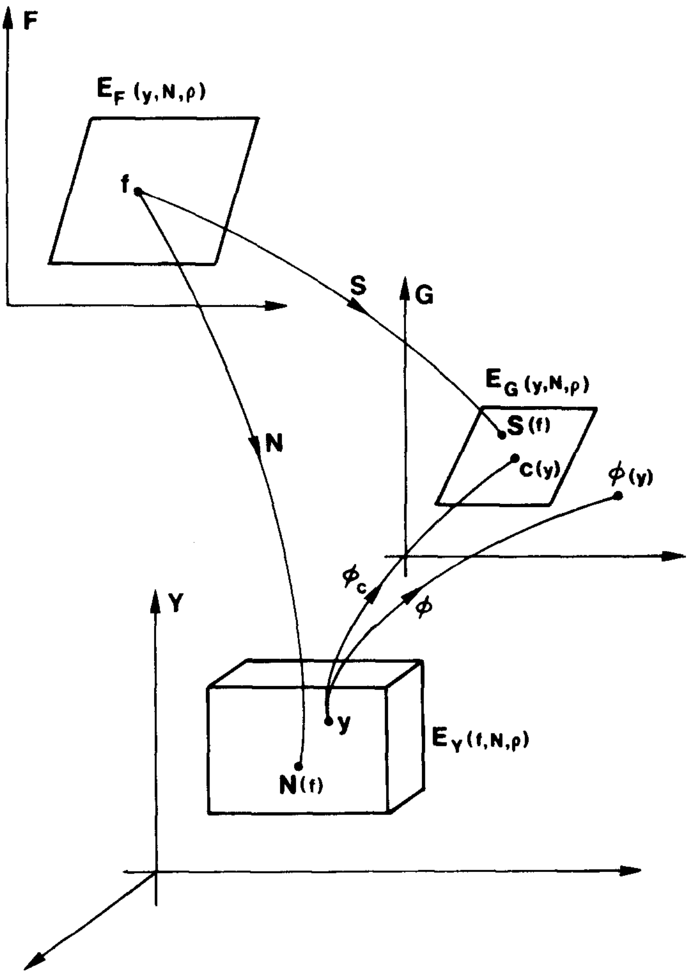

FIGURE 1

3. LOCAL ERRORS AND STRONGLY OPTIMAL ALGORITHMS

We define a local error $e_F(\phi, f, N, \rho)$ of approximation of $S(f)$ using the algorithm ϕ and the approximate information y for each problem element f belonging to the set F_0:

$$e_F(\phi, f, N, \rho) = \sup_{y \in E_Y(f, N, \rho)} \|S(f) - \phi(y)\|. \tag{6}$$

We call the f-local radius of approximate information with respect to a class of algorithms Φ the quantity

$$r_F^{\Phi}(f, N, \rho) = \inf_{\phi \in \Phi} e_F(\phi, f, N, \rho). \tag{7}$$

Remark that the above quantity may vanish if the class Φ is not suitably restricted to exclude nonfeasible algorithms; as a consequence, it will always be necessary to consider restricted classes of algorithms when dealing with local problems in the F space.

In a similar way we define a local error of approximation $e_Y(\phi, y, N, \rho)$ for each approximate information y and a y-local radius of approximate information in the Y space:

$$e_Y(\phi, y, N, \rho) = \sup_{f \in E_F(y, N, \rho)} \|S(f) - \phi(y)\| \tag{8}$$

$$r_Y^{\Phi}(y, N, \rho) = \inf_{\phi \in \Phi} e_Y(\phi, y, N, \rho). \tag{9}$$

Now we define strongly optimal algorithms in relation to the local problems introduced above. We can consider two different kinds of strong optimality related to the local errors (6) and (8).

An algorithm ϕ_f is called f-strongly optimal (for a worst case) in the class Φ if its local error e_F attains the local radius r_F, for each problem element f:

$$e_F(\phi_f, f, N, \rho) = r_F^{\Phi}(f, N, \rho) \qquad \forall f \in F_0. \tag{10}$$

An algorithm ϕ_y is called y-strongly optimal (for a worst case) in the class Φ if

$$e_Y(\phi_y, y, N, \rho) = r_Y^{\Phi}(y, N, \rho) \qquad \forall y \in Y_0. \tag{11}$$

An algorithm ϕ is called strongly optimal (for a worst case) if it is both f- and y-strongly optimal.

Similar but not exactly equivalent local errors and optimality concepts are also investigated by Traub *et al.* (1983) in a more general context, where F, Y, and G are assumed as general sets. It is important to notice that local errors as defined here are of particular interest in problems of system parameter, state estimation, or time series prediction. In fact, in these problems typically a set of measurements y is available and one must determine optimal estimates of $S(f)$ for each possible y using an algorithm $\phi(y)$. On the other hand, f-strong optimality is also a particularly meaningful property in estimation problems, as it ensures the minimum uncertainty of the estimates with regard to all possible measurements $y \in E_Y(f, N, \rho)$, for a particular though unknown element $f \in F_0$. Furthermore, f-strong optimality is relevant also from the point of view of optimal information problems. In fact, in this case the information y is not available a priori and the aim is the computation of an optimal information operator minimizing the local radius of information r_F^{Φ}. Actually, the local error e_F and the radius r_F^{Φ} cannot be computed exactly, since they depend on the unknown problem element f. Since the error e_Y can be computed in practical situations, relations between the local errors e_F and e_Y may be very useful and will be investigated in the following.

We now introduce global problems in a worst-case setting and optimal error algorithms for global problems. Such problems have already been considered by Micchelli and Rivlin (1977), Traub and Woźniakowski (1980), Traub *et al.* (1983), and Milanese and Tempo (1985). Let $e(\phi, N, \rho)$ be the (global) error:

$$e(\phi, N, \rho) = \sup_{f \in F_0} e_F(\phi, f, N, \rho) = \sup_{y \in Y_0} e_Y(\phi, y, N, \rho). \qquad (12)$$

Furthermore, let $r(N, \rho)$ be the (global) radius of information for Φ defined as the class of all mappings:

$$r(N, \rho) = \inf_{\phi \in \Phi} e(\phi, N, \rho). \qquad (13)$$

An algorithm ϕ_0 for which

$$e(\phi_0, N, \rho) = r(N, \rho) \qquad (14)$$

is called an optimal error algorithm. A general class of optimal algorithms, which has been used in estimation and prediction contexts (Milanese *et al.*, 1984; Milanese and Tempo, 1985) is that of central algorithms. Let $c(y)$ be a center of the set $E_G(y, N, \rho)$, i.e.,

$$\sup_{u \in E_G(y,N,\rho)} \|c(y) - u\| = \inf_{g \in G} \sup_{u \in E_G(y,N,\rho)} \|g - u\|. \qquad (15)$$

A central algorithm ϕ_c is such that

$$\phi_c(y) = c(y). \qquad (16)$$

Global optimality of central algorithms has been proved by Micchelli and Rivlin (1977). Actually, central algorithms enjoy properties even stronger than global optimality. In fact, it follows from definitions (15) and (16) that a central algorithm minimizes the local error $e_Y(\phi, y, N, \rho)$ for every approximate information y, i.e.,

$$e_Y(\phi_c, y, N, \rho) = r_Y(y, N, \rho) \qquad \forall y \in Y_0, \qquad (17)$$

which means that every central algorithm is a y-strongly optimal algorithm.

Another relevant class of algorithms, often considered in estimation problems, is the class $\Phi^C \subset \Phi$ of correct algorithms (Milanese and Belforte, 1982). An algorithm $\phi \in \Phi^C$ is a correct algorithm if

$$\phi(N(f)) = S(f) \qquad \forall f \in F_0. \tag{18}$$

Note that most of the usual classes of estimators (such as least squares, minimum absolute errors, etc.) are correct.

We denote by $\Phi^{CL} \subset \Phi^{C}$ the class of all correct and linear algorithms. It should be remarked that the concept of correct algorithm is meaningful only in estimation problems where $\dim F \leq \dim Y$. In these cases it is usually supposed that the information is complete (i.e., N is a one-to-one mapping) and the solution uncertainty is due only to the information error η. In fact, when exact information is given, the global radius is always zero.

4. Linear Estimation Problems

In both this and the following section, unless otherwise specified, we will assume that the solution operator S is linear and the information operator N is linear and complete (i.e., is a linear one-to-one mapping). The following theorem establishes connections between the local errors e_F and e_Y and the global error $e(\phi, N, \rho)$ in the class of linear and correct algorithms Φ^{CL}. In particular, it is shown that the local error e_F is equal to the global error for each problem element f and, at the same time, it is an upper bound of the local error e_Y.

THEOREM 1 (Milanese and Tempo, 1985). *If S, N, and ϕ are linear, ϕ is correct, and $F_0 = F$, then the following relationships hold:*

$$e_F(\phi, f, N, \rho) = e(\phi, N, \rho) \geq e_Y(\phi, y, N, \rho) \qquad \forall f \in F, \forall y \in Y_0. \tag{19}$$

The following lemma shows that linear optimal algorithms enjoy a stronger property than optimality, namely f-strong optimality within the class of linear and correct algorithms.

LEMMA 1. *Let S, N be linear operators and $F_0 \equiv F$. If ϕ_0 is an optimal, linear, and correct algorithm, then it is f-strongly optimal in the class Φ^{CL}.*

Proof. From relation (19) of Theorem 1 it follows that

$$e_F(\phi, f, N, \rho) = e(\phi, N, \rho) \qquad \forall \phi \in \Phi^{CL}; \tag{20}$$

then

$$\inf_{\phi \in \Phi^{CL}} e_F(\phi, f, N, \rho) = \inf_{\phi \in \Phi^{CL}} e(\phi, N, \rho). \tag{21}$$

Since ϕ_0 is optimal, then

$$e(\phi_0, N, \rho) = \inf_{\phi \in \Phi^{CL}} e(\phi, N, \rho). \tag{22}$$

Using definition (7) of f-local radius and relations (20) and (21) we obtain

$$e_F(\phi_0, f, N, \rho) = r_F^{\Phi^{CL}}(f, N, \rho) \qquad \forall f \in F. \tag{23}$$

■

The problem of existence of linear optimal algorithms has been studied by Marchuk and Osipenko (1975) when S is a linear functional and N is a linear and partial (i.e., it is not a one-to-one mapping) operator. This result has been extended to the case of linear operator S in Milanese and Tempo (1985). In the same paper a linear optimal and correct algorithm is derived in the case of a complete information operator under the following general condition on the spaces and operators involved.

Condition LP. F, Y, and G are n-, m-, and r-dimensional spaces, respectively, with $m \geq n$; G and Y are equipped with l_∞ and l_∞^W norms, respectively, and $F_0 \equiv F$.

As emphasized by Belforte *et. al.* (1982), Milanese *et al.* (1984), and Milanese and Tempo (1985), Condition LP is not restrictive in many of the application areas of estimation and prediction theory. On the other hand, such a condition is of particular interest because under its hypotheses both optimal linear algorithms and central algorithms can be easily computed by linear programming techniques. Furthermore, under this condition we are able to prove a further property of central algorithms which represents the main result of this section.

THEOREM 2. *If Condition* LP *holds there exists a central algorithm ϕ_c in the class Φ^C; moreover, such an algorithm is strongly optimal in the class Φ^C.*

Proof. The existence of a correct central algorithm follows from the fact that the central algorithm derived by Milanese and Tempo (1985) is correct.

In order to prove the second part of the theorem let us define the sets

$$E_F(N\hat{f}, N, \rho) = \{f \in F : \|Nf - N\hat{f}\| \leq \rho\} \tag{24}$$

$$E_Y(\hat{f}, N, \rho) = \{y \in Y_0 : \|y - N\hat{f}\| \leq \rho\}. \tag{25}$$

From definitions (24) and (25) it follows that

$$E_Y(\hat{f}, N, \rho) \supseteq N\{E_F(N\hat{f}, N, \rho)\} \qquad \forall \hat{f} \in F. \tag{26}$$

Define

$$E_G(N\hat{f}, N, \rho) = S\{E_F(N\hat{f}, N, \rho)\} \tag{27}$$

$$E_G(\phi, \hat{f}, N, \rho) = \phi\{E_Y(\hat{f}, N, \rho)\}. \tag{28}$$

For any correct algorithm $\phi \in \Phi^C$ we have

$$\phi\{N\{E_F(N\hat{f}, N, \rho)\}\} = E_G(N\hat{f}, N, \rho). \tag{29}$$

From (26) it easily follows that

$$\phi\{E_Y(\hat{f}, N, \rho)\} \supseteq E_G(N\hat{f}, N, \rho) \qquad \forall \hat{f} \in F, \tag{30}$$

which by definition (28) becomes

$$E_G(\phi, \hat{f}, N, \rho) \supseteq E_G(N\hat{f}, N, \rho) \qquad \forall \hat{f} \in F, \forall \phi \in \Phi^C. \tag{31}$$

Using definitions (27) and (28) we obtain

$$\sup_{f \in E_F(N\hat{f}, N, \rho)} \|S\hat{f} - Sf\| = \sup_{g \in E_G(N\hat{f}, N, \rho)} \|S\hat{f} - g\| \tag{32}$$

$$\sup_{y \in E_Y(\hat{f}, N, \rho)} \|S\hat{f} - \phi(y)\| = \sup_{g \in E_G(\phi, \hat{f}, N, \rho)} \|S\hat{f} - g\|. \tag{33}$$

From (31), (32), and (33) it follows that

$$\sup_{f \in E_F(N\hat{f}, N, \rho)} \|S\hat{f} - Sf\| \leq \sup_{y \in E_Y(\hat{f}, N, \rho)} \|S\hat{f} - \phi(y)\| \qquad \forall \hat{f} \in F, \forall \phi \in \Phi^C. \tag{34}$$

Since S and N are linear operators,

$$\sup_{f \in E_F(N\hat{f}, N, \rho)} \|S\hat{f} - Sf\| = \sup_{f: \|N(f - \hat{f})\| \leq \rho} \|S(\hat{f} - f)\| = \sup_{h: \|Nh\| \leq \rho} \|Sh\|. \tag{35}$$

The global radius $r(N, \rho)$ under Condition LP is given by

$$r(N, \rho) = \sup_{h: \|Nh\|_\infty^W \leq \rho} \|Sh\|_\infty. \tag{36}$$

Since a central algorithm ϕ_c is an optimal algorithm, using definition (13) we get

$$r(N, \rho) = \sup_{\hat{f} \in F} \sup_{y \in E_Y(\hat{f}, N, \rho)} \|S\hat{f} - \phi_c(y)\| \geq \sup_{y \in E_Y(\hat{f}, N, \rho)} \|S\hat{f} - \phi_c(y)\|. \tag{37}$$

From (34), (35), (36), and (37) it follows that

$$\sup_{y \in E_Y(\hat{f}, N, \rho)} \|S\hat{f} - \phi_c(y)\|_\infty \leq \sup_{y \in E_Y(\hat{f}, N, \rho)} \|S\hat{f} - \phi(y)\|_\infty \qquad \forall \hat{f} \in F, \forall \phi \in \Phi^C, \tag{38}$$

which proves f-strong optimality of central algorithms in the class Φ^C. Since central algorithms are also y-strongly optimal, the proof is complete. ∎

Theorem 2 states that a central algorithm ϕ_c minimizes the local error e_F for every problem element f in the class of correct algorithms. This result and (17) show that under Condition LP a central algorithm minimizes both local errors e_Y and e_F. Let us now turn our attention to the meaning of such properties in estimation problems. The error e_Y represents the estimation error obtained with a fixed set of measurements, corresponding to the particular observed realization of the experiment examined, and is due to the fact that a whole set of problem elements f are possible candidates to represent the true problem element. On the other hand, e_F, as already mentioned, is the estimation error due to all possible outcomes of the experiment compatible with the maximum assumed uncertainty, once the model f has been fixed. The property of minimizing e_Y (for each $y \in Y_0$) and e_F (for each $f \in F$) is a major requisite for a valuable estimator; in fact, it guarantees the robustness of the algorithm with respect to all possible variations of the data or of the problem element due to the admissible intrinsic uncertainty of the problem setting.

5. Linear Optimal Information in Estimation Problems

5.1. *Adaptive and Nonadaptive Information*

We consider some questions related to optimal information. In dealing with this problem the approximate information y is not assumed as given a priori and we are concerned with global and f-local problems.

In the classical theory of optimal algorithms two different classes of information are considered: adaptive and nonadaptive information (Traub and Woźniakowski, 1980; Traub *et al.*, 1983).

A nonadaptive information operator is defined as

$$N^{\text{non}}(f) = \{L_1(f), L_2(f), \ldots, L_m(f)\}, \tag{39}$$

where $L_1, \ldots, L_m$ are linear functionals; the number m is called cardinality of the information and is denoted by $\text{card}(N^{\text{non}})$.

Two kinds of adaptive information may be defined,

$$N^{\text{a}}(f) = \{L_1(f), L_2(f; L_1(f)), \ldots, L_m(f; L_1(f), \ldots, L_{m-1}(f))\} \tag{40}$$

$$N^{\text{a}}(f; y) = \{L_1(f), L_2(f; y_1), \ldots, L_m(f; y_1, \ldots, y_{m-1})\}, \tag{41}$$

where $L_1, \ldots, L_m$ are linear functionals in f and $y_1, \ldots, y_{m-1}$ are values of the approximate information; as for N^{non}, the cardinality of N^{a} will be denoted by $\text{card}(N^{\text{a}})$. The linear functionals L_i in (40) may depend on the previous exact values $L_1, \ldots, L_{i-1}$, while in (41) they are allowed to depend on the previous approximate information instead of the exact values.

Although in the context of approximate information the adaptive information $N^a(f; y)$ may be more relevant from a practical point of view, the information $N^a(f)$ is expected to be more powerful; for this reason in the following we will investigate the relations between the radii of information of $N^a(f)$ and $N^{non}(f)$.

In general it is clear from definitions (39) and (40) that the structure of an adaptive information operator is more general than that of a nonadaptive one. So, if we denote by $\Psi_m^a(\Psi_m^{non})$ the class of linear adaptive (nonadaptive) information operators with cardinality less or equal to m, we have

$$\Psi_m^a \supset \Psi_m^{non}. \tag{42}$$

An optimal adaptive (nonadaptive) information is defined as an operator N_0^a (N_0^{non}) which minimizes the global radius within a fixed class Ψ_m^a (Ψ_m^{non}):

$$\begin{aligned} r(N_0^a, \rho) &= \inf_{N \in \Psi_m^a} r(N, \rho) \\ (r(N_0^{non}, \rho) &= \inf_{N \in \Psi_m^{non}} r(N, \rho)). \end{aligned} \tag{43}$$

Analogous definitions of f-local (in the class Φ) optimal information are obtained by substitution of the f-local radius r_F in place of the global radius.

In the following we give some results on the relations between the radii of adaptive and nonadaptive information in estimation contexts.

THEOREM 3 (Traub *et al.*, 1983). *If S is linear, Y is a linear normed space, and F_0 is a balanced and convex set, then*

$$r(N_0^a, \rho) \leq r(N_0^{non}, \rho) \leq 2r(N_0^a, \rho). \tag{44}$$

Theorem 3 states that in a general linear context, adaptive information is no more effective than nonadaptive information within a constant of two. We show that for problems satisfying Condition LP, adaptive information is as powerful as nonadaptive information both for global and for f-local (in class Φ^{CL}) problems.

THEOREM 4. *If Condition* LP *holds and S is linear, then*

$$r(N_0^a, \rho) = r(N_0^{non}, \rho) \tag{45}$$

$$r_F^{\Phi^{CL}}(f, N_0^a, \rho) = r_F^{\Phi^{CL}}(f, N_0^{non}, \rho) \qquad \forall f \in F. \tag{46}$$

Proof. From relation (19) of Theorem 1 it follows that

$$e_F(\phi, f, N, \rho) = e(\phi, N, \rho) \qquad \forall f \in F, \forall \phi \in \Phi^{CL} \tag{47}$$

and, consequently, using definition (7),

$$\inf_{\phi \in \Phi^{\mathrm{CL}}} e(\phi, N, \rho) = r_F^{\Phi^{\mathrm{CL}}}(f, N, \rho) \qquad \forall f \in F. \tag{48}$$

If Condition LP holds then there exists an optimal algorithm in the class Φ^{CL} (Milanese and Tempo, 1985) and therefore from (48) and (13),

$$r(N, \rho) = \inf_{\phi \in \Phi^{\mathrm{CL}}} e(\phi, N, \rho) = r_F^{\Phi^{\mathrm{CL}}}(f, N, \rho) \qquad \forall f \in F. \tag{49}$$

It must be remarked that (49) holds true for both adaptive and nonadaptive information.

The remaining part of the proof follows the line of Theorem 4.1 of Traub *et. al.* (1983, p. 63), which refers to a slightly different definition of local errors. Define the following information operator for a fixed problem element h:

$$N_h^{\mathrm{non}}(f) = \{L_1(f), L_2(f; L_1(h)), \ldots, L_m(f; L_1(h), \ldots, L_{m-1}(h))\}. \tag{50}$$

For $f = h$ we get

$$N_h^{\mathrm{non}}(h) = N^{\mathrm{a}}(h). \tag{51}$$

From relations (49) and (51) we obtain that

$$r(N_h^{\mathrm{non}}, \rho) = r_F^{\Phi^{\mathrm{CL}}}(h, N_h^{\mathrm{non}}, \rho) = r_F^{\phi^{\mathrm{CL}}}(h, N^{\mathrm{a}}, \rho) = r(N^{\mathrm{a}}, \rho). \tag{52}$$

Since for every $N \in \Psi_m^{\mathrm{a}}$ there exists $h \in F$ for which (51) holds, we can take the infimum in (52), obtaining

$$\inf_{N \in \Psi_m^{\mathrm{non}}} r(N, \rho) \leq \inf_h r(N_h^{\mathrm{non}}, \rho) = \inf_{N \in \Psi_m^{\mathrm{a}}} r(N, \rho). \tag{53}$$

By taking into account relation (42), the first statement of the theorem follows.

The second statement of the theorem simply follows from (49). ■

Since Theorem 4 shows that, under Condition LP, adaptation in the sense of definition (40) does not help, it easily follows also that adaptive information depending on the approximate values according to definition (41) does not help. Therefore from now on we shall not distinguish between the classes Ψ_m^{a} and Ψ_m^{non} and we will drop the superscripts "a" and "non."

5.2. *Optimal Sampling Design for Estimation Problems*

The problem of optimal sampling design has been studied in depth in recent years either in a statistical setting, where the noise is described by a suitable statistical distribution (Fedorov, 1972; Goodwin and Payne, 1977; Mori and Di Stefano, 1979) or in a deterministic setting, where the measurements are assumed corrupted by unknown but bounded additive noise (Belforte *et al.*, 1984). The problem consists in looking for optimal sampling times of a time function $f(t)$ over a given time interval which allow one to minimize some given criteria.

The optimal sampling design can be reduced to a linear optimal information problem with particular choices of F and Ψ_m. In fact suppose $f \in F$ as a function of time $f(t)$, and consider only information operators defined as sampling operators,

$$N(f) = \{f(t_1), f(t_2), \ldots, f(t_m)\}, \tag{54}$$

where the sampling times t_i belong to a discrete set $\{\tau_1, \tau_2, \ldots, \tau_M\}$ with $M > m$ and $\tau_i \neq \tau_j$, $i \neq j$. Then any particular N of cardinality m is defined by choosing a subset made of m elements out of $\{\tau_1, \tau_2, \ldots, \tau_M\}$. With a slight abuse of notation N will denote both an information operator and the corresponding sampling times $t_1, t_2, \ldots, t_m$.

Denoting by Ψ_m^s the class of all sampling operators of cardinality less or equal to m, the optimal sampling design (Belforte *et al.*, 1984) consists in looking for f-locally optimal information in the class Φ^{CL}, restricted to class Ψ_m^s; this means that $N_0 \in \Psi_m^s$ is called an optimal sampling operator if

$$r_F^{\Phi^{CL}}(f, N_0, \rho) = \inf_{N \in \Psi_m^s} r_F^{\Phi^{CL}}(f, N, \rho) \qquad \forall f \in F. \tag{55}$$

Let us assume that Condition LP holds and consider the information operator N^τ of cardinality M:

$$N^\tau(f) = \{f(\tau_1), f(\tau_2), \ldots, f(\tau_M)\}. \tag{56}$$

Furthermore, let us define information operators N^i of cardinality n such that

$$\sup_{\|N^i(f)\|_\infty^w = \rho} (Sf)_i = \sup_{\|N^\tau(f)\|_\infty^w \leq \rho} (Sf)_i, \qquad i = 1, \ldots, r. \tag{57}$$

Note that the existence of at least one N^i is due to the fact that each problem in (57) is a linear programming problem in an n-dimensional space. The equality constraints in (57), called active constraints, define the operators N^i. Uniqueness is not guaranteed, but any N^i satisfying (57) may be used.

Let us now define the information operator $\overline{N}$ as follows:

$$\overline{N} = \bigcup_{i=1}^{r} N^i. \tag{58}$$

This means that $\bar{N}$ has as sampling times the union of sampling times of all N^i and that:

$$n \leq \operatorname{card}(\bar{N}) \leq n \cdot r. \tag{59}$$

The following theorem, which is a generalization of a result presented by Belforte *et al.* (1984), shows that for problems satisfying Condition LP the optimal sampling design may be solved by computing the active constraints of suitable linear programming problems.

THEOREM 5. *Let Condition* LP *hold. If* $\operatorname{card}(\bar{N}) \leq m$ *then*

(i) $\bar{N}$ *is an optimal sampling operator for f-local (in the class* Φ^{CL}*) problems:*

$$r_F^{\Phi^{\mathrm{CL}}}(f, \bar{N}, \rho) = \inf_{N \in \Psi_m^s} r_F^{\Phi^{\mathrm{CL}}}(f, N, \rho) \qquad \forall f \in F. \tag{60}$$

(ii)
$$r_F^{\Phi^{\mathrm{CL}}}(f, \bar{N}, \rho) = \sup_i \sup_{f \in F : \|N^i(f)\|_\infty^w \leq \rho} (Sf)_i. \tag{61}$$

Proof. From Theorem 1 it follows that

$$\inf_{\phi \in \Phi^{\mathrm{CL}}} e(\phi, N, \rho) = \inf_{\phi \in \Phi^{\mathrm{CL}}} e_F(\phi, f, N, \rho) \qquad \forall N \in \Psi. \tag{62}$$

If Condition LP holds then there exists an optimal algorithm in the class Φ^{CL} (Milanese and Tempo, 1985) and therefore

$$r(N, \rho) = \inf_{\phi \in \Phi^{\mathrm{CL}}} e(\phi, N, \rho) \qquad \forall N \in \Psi. \tag{63}$$

Using definition (7) and relations (62) and (63) we obtain

$$r(\bar{N}, \rho) = r_F^{\Phi^{\mathrm{CL}}}(f, \bar{N}, \rho). \tag{64}$$

Since Condition LP holds and the set $E_F(0, \bar{N}, \rho)$ is balanced, the local radius can be expressed as

$$r_F^{\Phi^{\mathrm{CL}}}(f, \bar{N}, \rho) = r(\bar{N}, \rho) = \sup_{f \in E_F(0, \bar{N}, \rho)} \sup_i (Sf)_i = \sup_i \sup_{f \in E_F(0, \bar{N}, \rho)} (Sf)_i. \tag{65}$$

Using the definition (58) we get

$$\sup_{f \in E_F(0, \bar{N}, \rho)} (Sf)_i = \sup_{\|\bar{N}(f)\|_\infty^w \leq \rho} (Sf)_i = \sup_{\|N^i(f)\|_\infty^w = \rho} (Sf)_i, \qquad i = 1, \ldots, r \tag{66}$$

and from (65) and (66), (61) follows.

From the definitions of $\bar{N}$ and N^i, using relations (57) and (58) it directly follows that

$$r_F^{\Phi CL}(f, \overline{N}, \rho) = r_F^{\Phi CL}(f, N^\tau, \rho) \qquad \forall f \in F, \tag{67}$$

Since $N \in \Psi_m^s$ is a subset of N^τ then

$$r_F^{\Phi CL}(f, N^\tau, \rho) \leq r_F^{\Phi CL}(f, N, \rho) \qquad \forall N \in \Psi_m^s, \forall f \in F, \tag{68}$$

which completes the proof. ■

Remark 1. Since from (63) and (64) it easily follows that

$$r(N, \rho) = r_F^{\Phi CL}(f, N, \rho) \qquad \forall N \in \Psi, \tag{69}$$

it results that the information operator $\overline{N}$ is also globally optimal within the class Ψ_m^s and its radius $r(\overline{N}, \rho)$ attains the local radius $r_F^{\phi CL}(f, \overline{N}, \rho)$.

Remark 2. Theorem 5 gives a solution of the optimal sampling design only if $\text{card}(\overline{N}) \leq m$. From (59) it follows that if $\text{card}(\overline{N}) > n$ this condition may not be met even if $m \geq n$ (which is usual in the estimation and prediction fields). In Belforte *et al.* (1984) examples are shown in which $\text{card}(\overline{N}) > n$. In the same paper it is shown that there exist some possible spaces F (for example, the space of polynomials of degree $n - 1$) such that $\text{card}(\overline{N}) = n$; furthermore, in this case $\overline{N}$ results to be nonsingular assuming that N^τ is complete, which, as previously mentioned, is a reasonable and usual assumption in estimation problems. In such a case a linear central algorithm which uses the optimal sampling $\overline{N}$ can be easily computed as shown by the following theorem.

THEOREM 6. *Let Condition* LP *hold. If* $\text{card}(\overline{N}) = n$, *a linear central algorithm* (*using approximate optimal information* $\overline{y} = \overline{N}f + \overline{\eta}$) *is given by*

$$\phi^*\overline{y} = S\overline{N}^{-1}\overline{y}. \tag{70}$$

Proof. Let us consider an arbitrary approximate information $\overline{y}$; under the invertibility condition on $\overline{N}$ we can define an element $\overline{f} \in E_F(\overline{y}, \overline{N}, \rho)$ such that

$$\overline{f} = \overline{N}^{-1}\overline{y}. \tag{71}$$

Now we show that $\overline{f}$ is a symmetry center of the set $E_F(\overline{y}, \overline{N}, \rho)$. Consider an arbitrary element $f_1 \in E_F$ and an element f_2 defined as

$$f_2 = 2\overline{f} - f_1. \tag{72}$$

Since $f_1 \in E_F$,

$$\|\overline{N}f_1 - \overline{y}\| \le \rho. \tag{73}$$

Substituting (71) and (72) in (73) we obtain

$$\|\overline{N}f_2 - \overline{y}\| \le \rho, \tag{74}$$

which shows that f is a symmetry center of E_F.

Define the element $\overline{g} = S\bar{f} \in E_G(\overline{y}, \overline{N}, \rho)$; since S is linear it can be easily shown, as for $\bar{f}$, that $\overline{g}$ is a symmetry center of E_G.

Therefore the algorithm $\phi^* = S\overline{N}^{-1}$ takes an arbitrary $\overline{y}$ to a symmetry center of E_G, which proves that ϕ^* is a central algorithm. ■

In the following we give two examples, referring to a space F made of polynomial functions, where an optimal sampling operator and the corresponding coefficients of the algorithm (70) are computed.

EXAMPLE 1. Let F be the space of linear functions

$$F = \{f(t) : f(t) = a + bt\},$$

where a and b are unknown real coefficients. Consider the solution operator

$$Sf = \{a, b\}$$

and the information operator N^τ,

$$N^\tau f = \{f(\tau_1), \ldots, f(\tau_{100})\},$$

where $\tau_i = 0.01 \cdot i$, $i = 1100$. Assuming $\rho = 1$, $w_i = 1$, and $2 \le m < 100$ the following results are obtained:

$$\overline{N}f = \{f(0.01), f(1.00)\}$$

$$\phi^* = S\overline{N}^{-1} = \begin{bmatrix} 1.0101 & -0.0101 \\ -1.0102 & 1.0101 \end{bmatrix}$$

$$r(\overline{N}, \rho) = 2.0202$$

EXAMPLE 2. We consider the same problem as in the preceding example with $3 \le m < 100$ and the two following differences:

$$F = \{f(t) : f(t) = a + bt + ct^2\}$$

$$Sf = \{a, b, c\}.$$

The following results are obtained:

$$\overline{N}f = \{f(0.01), f(0.51), f(1.00)\}$$

$$\phi^* = S\overline{N}^{-1} = \begin{bmatrix} 1.0303 & -0.0408 & 0.0105 \\ 3.0505 & 4.1224 & -1.0719 \\ 2.0202 & -4.0816 & 2.0614 \end{bmatrix}$$

$$r(\overline{N}, \rho) = 8.2449.$$

References

Belforte, G., Bona, B., and Frediani, S., (1984), Optimal sampling schedule with unknown but bounded measurement errors: Families of linear models, *in* "Proceedings, 23rd IEEE Conference on Decision and Control, Las Vegas, Nevada."

Belforte, G., Milanese, M., and Tempo, R., (1982), Optimal algorithms theory and estimation with unknown but bounded errors, *in* "Proceedings, 21st IEEE Conference on Decision and Control, Orlando, Florida."

Fedorov, V. V. (1972), "Theory of Optimal Experiments" (W. J. Studden and E. M. Klinko, Trans. and Eds.), Academic Press, New York.

Goodwin, G. C., and Payne, R. L. (1977), "Dynamic System Identification: Experiment Design and Data Analysis," Academic Press, New York.

Marchuk, A. G., and Osipenko, K. Yu. (1975), Best approximation of functions specified with an error at a finite number of points, *Mat. Zametki* **17**, 359–368 (English transl.: *Math. Notes* **17** (1975), 207–212.

Micchelli, C. A., and Rivlin, T. J. (1977), A survey of optimal recovery, *in* "Optimal Estimation in Approximation Theory" (C. A. Micchelli and T. J. Rivlin, Eds.), Plenun, New York.

Milanese, M., and Belforte, G. (1982), Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: Linear families of models and estimators, *IEEE Trans. Automat. Control* **AC-27**, 408–414.

Milanese, M. and Tempo, R. (1985), Optimal algorithms theory for robust estimation and prediction, *IEEE Trans. Automat. Control* **AC-30**, 730–738.

Milanese, M., Tempo, R., and Vicino, A. (1984), Robust time series prediction by optimal algorithms theory, *in* "Proceedings, 23rd IEEE Conference on Decision and Control, Las Vegas, Nevada."

Mori, F., and Di Stefano, J. J. (1979), Optimal nonuniform sampling interval and test-input design for identification of physiological systems from very limited data, *IEEE Trans. Automat. Control* **AC-24**, 893–900.

Traub, J. F., Wasilkowski, G., and Woźniakowski, H. (1983), "Information, Uncertainty, Complexity," Addison–Wesley, Reading, Mass.

Traub, J. F., and Woźniakowski, H. (1980), "A General Theory of Optimal Algorithms," Academic Press, New York.